ORIGINAL PAPER

DOI: 10.26794/2587-5671-2024-28-5-146-155
JEL C15, C12, C13, G52

Statistical Analysis of Stable Distribution Application in Non Life Insurance

A. Laouar[a], K. Boukhetala[b], R. Sabre[c]
[a, b] University of science and technology Houari Boumedienne (USTHB), Algiers, Algeria;
[c] National Higher Institute of Agronomic, Food and Environmental Sciences, AgroSup Dijon, Dijon, France

ABSTRACT

In recent years, the theory of stable variables has seen many exciting developments, due to the fact that it is a very rich class of probability laws able to represent different asymmetries, and heavy tails, so modelling complex phenomena; unlike normal law, which very often underestimates extreme events. α-stable distributions are a class of heavy-tailed distributions. For that, we will start in this paper by presenting a review of graphical tests, which will help us to verify if we are in the presence of data with infinite variance or not, and more precisely of stable distribution. Then we will apply these tests to real data representing car claim amounts, allowing us to assume that our sample follows a stable distribution. In order to confirm this hypothesis, we will therefore estimate the four parameters of the distribution using the McCuloch method, as well as the Koutrouvelis method in order to be able to make the diagnosis with Kernel Densities, and finally we will demonstrate that α-stable distribution is better fitted to the car claim amount data by using the Kolmogorov test.

Keywords: stable distribution; infinite variance; simulation; statistical test

For citation: **Laouar A., Boukhetala K., Sabre R. Statistical analysis of stable distribution application in non life insurance. *Finance: Theory and Practice.* 2024;28(5):146-155. DOI: 10.26794/2587-5671-2024-28-5-146-155**

ОРИГИНАЛЬНАЯ СТАТЬЯ

Статистический анализ устойчивого распределения в страховании, кроме страхования жизни

А. Лауар[a], К. Бухетала[b], Р. Сабре[c]
[a, b] Университет науки и технологии Уари Бумедьена (USTHB), Алжир, Алжир;
[c] Национальный высший институт агрономических, пищевых и экологических наук (AgroSup), Дижон, Франция

АННОТАЦИЯ

В последние годы теория стабильных переменных претерпела множество захватывающих изменений, благодаря тому, что это связано с законом вероятности, представляющего различные асимметрии и статистику с «тяжелыми хвостами», что позволяет моделировать сложные явления в отличие от стандартного закона, который очень часто недооценивает экстремальные события. α-стабильные распределения — это класс распределений с "тяжелыми хвостами". В данной статье мы начнем с обзора графических тестов, которые помогут нам проверить, имеем ли мы данные с бесконечной дисперсией или нет, а точнее, стабильное распределение. Затем мы применим эти тесты к реальным данным, представляющим суммы страховых выплат по автомобилям, что позволит нам предположить, что наша выборка данных соответствует устойчивому распределению. Для подтверждения этой гипотезы мы оценим четыре параметра распределения с помощью метода МакКалоха, а также метода Кутрувелиса, чтобы иметь возможность провести диагностику с помощью плотности ядра, и, наконец, продемонстрируем, что α-устойчивое распределение лучше подходит для страховых выплат по автомобилям, используя тест Колмогорова.

Ключевые слова: стабильное распределение; бесконечная дисперсия; моделирование; статистический тест

Для цитирования: **Laouar A., Boukhetala K., Sabre R. Statistical analysis of stable distribution application in non life insurance. *Финансы: теория и практика.* 2024;28(5):146-155. DOI: 10.26794/2587-5671-2024-28-5-146-155**

INTRODUCTION

For any company, it's often essential to develop adequate strategies for efficient portfolio management to deal with probable risks. To do this, it is often sought to develop representative mathematical models, to be used as tools for analysis, forecasting and simulation for decision support. The choice of model is very important and as we know, Gaussian processes and variables have been studied for a long time and their usefulness in stochastic and statistical modelling is well accepted. However, they don't allow for large fluctuations and may sometimes be inadequate for modelling high variability. That's why it's important to focus on other families of laws and processes, such as stable random variables and processes, which naturally appear as alternative modeling tools. In recent years, the theory of stable variables has seen many exciting developments, due to the fact that it is a very rich class of probability laws able to represent different asymmetries, and heavy tails, so modelling complex phenomena.

α-stable distributions are a class of heavy-tailed distributions, this class was characterized by [1], in his paper the sum of independent and identically distributed variables. This class has a great importance in the theory of extreme values, because stable distributions can be characterized from the Generalized Central Limit Theorem given by Gnedenko and Kolmogorov (1954) [2] and indicates that if the condition of finite variance is not respected, the only possible limit law of the sum of n random variables (*iid*) is a Stable law. For all these reasons, we have chosen to focus on stable distribution for fitting claims amounts of car insurance.

HEAVY-TAILS DISTRIBUTION

In this section, we present briefly the notion of heavy-tailed distribution and various classes of such distributions. It is not easy to define heavy tails distribution precisely, but several definitions have been associated with such distributions according to classification criteria. The easiest characterization is based on the comparison with the normal law [3]. The distribution of a *r.v* X with mean M and variance σ^2 and is said to have a heavy tail if:

$$\frac{E\left[(X-M)^4\right]}{\sigma^4} > 3. \tag{1}$$

This is equivalent to saying that a distribution is heavy-tailed if and only if its kurtosis (the 4th central moment) is higher than the normal distribution (for which it's equal to 3), this indicates high peaks and fat tails (leptokurtic). Kurtosis less than three (<3) indicates lower peaks. The criterion, given by equation (1) is very general and can't be applied if the 4th moment of a random variable doesn't exist.

Unfortunately, it is not easy to define heavy tails precisely, and there is no criterion to classify all distributions relative to the right tail. We present in this section five classes of heavy-tailed distributions, borrowed from [4]:

- Distributions with no exponential moments (E);
- Subexponential distributions (D);
- Distributions with regular variations (C);
- Distributions with Pareto behaviour (B);
- α-Stable distributions with $\alpha < 2$ (A).

For these classes we have the following relationships:

$$A \subset B \subset C \subset D \subset E$$

These classes of distributions are nested, the broadest class E encompasses all distributions with $E\left(e^X\right) = \infty$. All distributions of class E are heavy-tailed with respect to the normal distribution (the tail probability $P(X > x) = 1 - F(x)$ of the normal distribution declines faster than exponentially).

ALPHA-STABLE DISTRIBUTION

The class of stable distributions is defined by means of their characteristic functions. With very few exceptions, no closed-form expressions are known for their densities and cumulative distribution functions, see [5] or [6, 7].

Definition: A random variable X is said to have a stable distribution, $X \in S_\alpha(\beta,\mu,\gamma)$ if its characteristic function $\phi_X(t) = Ee^{itX}$ has the following form:

$$\phi_X(t) = \exp\left\{i\mu t - \gamma^\alpha |t|^\alpha \left(1 - i\beta sign(t) W(\alpha,t)\right)\right\},\ t \in IR, \tag{2}$$

where

$$W(\alpha,t) = \begin{cases} tan\frac{\alpha\pi}{2} & if \quad \alpha = 1 \\ -\frac{2}{\pi} log|t| & if \quad \alpha \neq 1 \end{cases}$$

and

$$sign(t) = \begin{cases} 1 & if \quad t > 0 \\ 0 & if \quad t = 0 \\ -1 & if \quad t < 0 \end{cases}$$

The stable laws are described by four parameters:

- *Index of stability* $0 < \alpha \le 2$: determines the rate at which the tails of the distribution taper off. When $\alpha = 2$, the Gaussian distribution results, when $\alpha < 2$, the variance is infinite and the tails are asymptotically equivalent to a Pareto.
- *Skewness parameter* $\beta \in [-1,1]$: when β is positive (negative), the distribution is skewed to the right (left),

when $\beta = 0$, the distribution is symmetric about the location parameter μ. As α approach 2, β loses its effect and the distribution approaches the Gaussian distribution regardless of β.

- *Location parameter* $\mu \in IR$: determines the shift of the mode (the peak) of the density.
- *Scale parameter* $\gamma > 0$: determines the width, when $\gamma = 1$ and $\mu = 0$ the distribution is called standard stable.

STATISTICAL TESTS FOR STABLE LAW

In this section, we describe two graphics tests that may allow us to know if we are in the presence of an infinite variance law or not [8, 9]. For this, we suppose that we have a sequence of observations $(x_1, \dots, x_n)$.

Test 1:

This first test is the simplest and most used, it's decomposed into two parts:

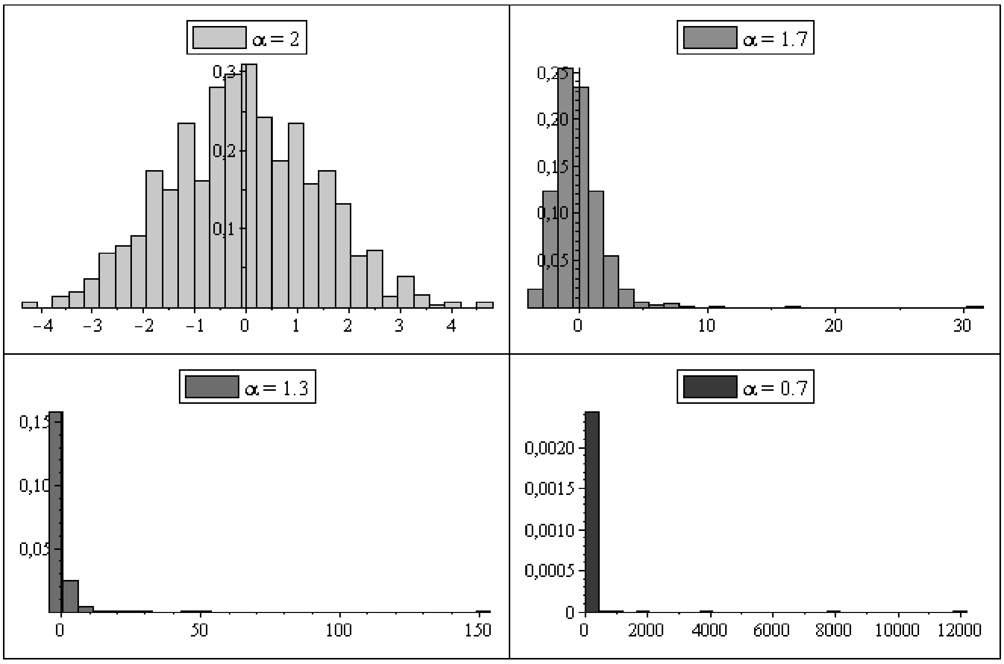

Fig. 1. **Histogram of** $S_\alpha(1,0,1)$

Source: Compiled by the authors.

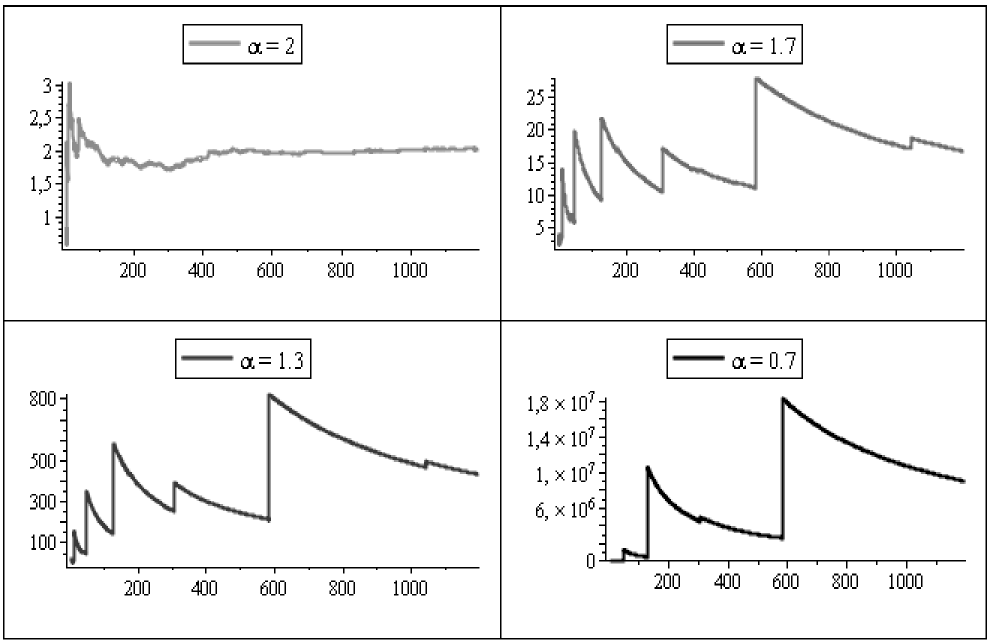

Fig. 2. **Test 1 for** $S_\alpha(1,0,1)$ ***r.v***

Source: Compiled by the authors.

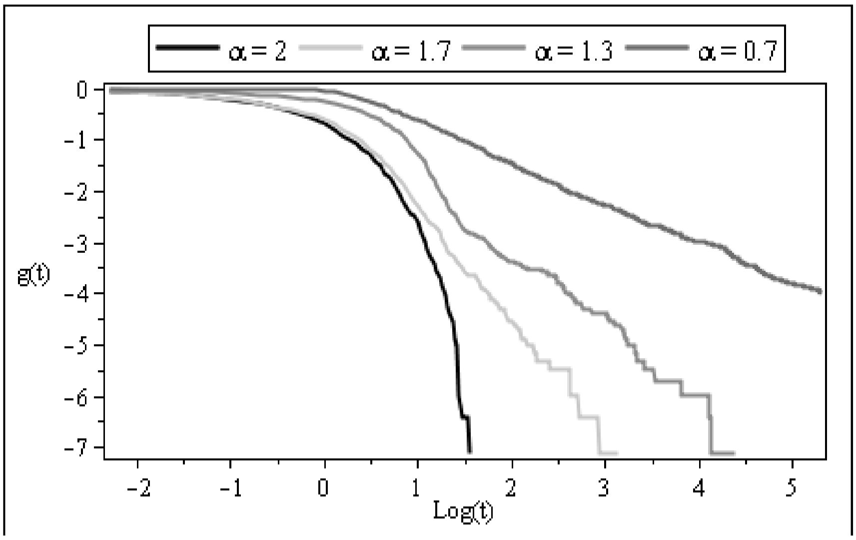

Fig. 3. **Test 2 of $S_\alpha(1,0,1)$ *r.v* for Different Value of α**

Source: Compiled by the authors.

- Calculate the variance for different values of n:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \quad (3)$$

- Draw the graph (n, S_n^2)

Distribution has finite variance, then there exists a finite constant c, such as:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \to c \text{ as } n \to \infty \text{ Almost surely,}$$

and vice versa.

When n increases and when the variance is finite, the plot must converge (see *Fig. 2* for $\alpha = 2$); On the contrary, if we are in the presence of a law with infinite variance, the plot diverges, and does not especially grow exponentially, as some received ideas suggest.

Test 2:

This second test is based on the fact that stable distributions have asymptotically the same behaviour as a Pareto distribution:

$$\lim_{t\to\infty} t^\alpha P(|X| > t) = \gamma C(\alpha) \quad (4)$$

So in $+\infty$ we have $\frac{d \log P(|X|) > t}{d \log t}$

$\frac{d \log IP(|X| > t)}{d \log t}$ $\frac{d \log IP(|X| > t)}{d \log t}$ is equivalent to α

Also here we have two steps:

- Fix t and calculate $g(t) = log\left(\frac{1}{n}\sum_{i=1}^{n} 1_{|x_i|>t}\right)$
- Draw the graph $(\log t, g(t))$ and see if the slope is finite from some value of t.

EXAMPLES OF GRAPHICS TESTS

For understanding the previous tests, we simulated by Chambers, Mallows and stuck methods [10, 11] many sequences of α-stable *r.v* for different values of α : 0.7, 1.3, 1.7 and $\alpha = 2$, with $\beta = 1$, $\mu = 0$ and $\gamma = 1$. The r.v. is $S_\alpha(1,0,1)$, for this, the distribution is skewed to the right. In first, we present the histogram of stable distribution for different values of α in *Fig. 1*.

As shown in *Fig. 2*, for a population with finite variance (like Gaussian distribution for $\alpha = 2$), the partial variance soon settles down close to the population variance (converges to a constant). For a population with an infinite variance, we see jumps up in the partial variance followed by slow declines until the next very large value appears in the sample. When n increases, the series of empirical variance not only diverges, but also oscillates with a high frequency for < 2. But when $\alpha = 2$ the series of variances no longer varies and becomes stable, and in *Fig. 3*, only the case $\alpha = 2$ gives a finite slope (the slope is a vertical line when $n \to \infty$).

APPLICATION IN NON-LIFE INSURANCE

Standard Mathematical Model

In the most general case a risk process $R(t)$ representing the behaviour of an insurance company is described by the following equation:

$$R(t) = u + ct - S(t). \quad (5)$$

Where u is the initial capital, c is the constant premium rate and $S(t)$ is the claim process defined by:

$$S(t)=\sum_{i=1}^{N(t)} X_i \tag{6}$$

$\{X_i\}_{i\geq 1}$ Sequence of independent, positive, identically distributed *r.v*, which represent claim severities.

N_t : Number of claims in (0, *t*], we assume that X_t and N_t are independent.

This model is known as the classical risk process or Cramér-Lundberg model [12], where $S(t)$ is a compound Poisson.

As we can see, the aggregate claim amount $S(t)$ is a random sum of random variables. And for good risk management of the insurance company, it needed to have good modelling of this sum (claim process), which depends essentially of claims amounts and their frequencies. To do this, we must examine sequences of real data to have the best estimate of the claim amount distribution.

Statistical Analysis of Real Data

For a good modelling of the risk, it's important to know the distribution of claim amounts. For this, we propose to study the daily real data from an insurance company over a period of one year (2017 and 2018) using Matlab R 2021a. *Fig. 4* and *Fig. 5* give us the behaviours of claims amounts.

It's clear that there is considerable jump in the claims amount, especially in 2018.

In *Fig. 6*, we can see that we have the asymmetric leptokurtic features, that is, the claim distribution is skewed to the right, and has a higher peak and heavier tail than those of the normal distribution.

Histograms of our data are similar to the histogram of α-stable distribution for $\alpha < 2$ (see *Fig. 1*), which is skewed right. So, the α-stable distribution can be seen as a useful tool to capture the asymmetric leptokurtic features of the claim amount, which is confirmed in *Fig. 7* and *Fig. 8* corresponding to the results of tests 1 and 2 for real data.

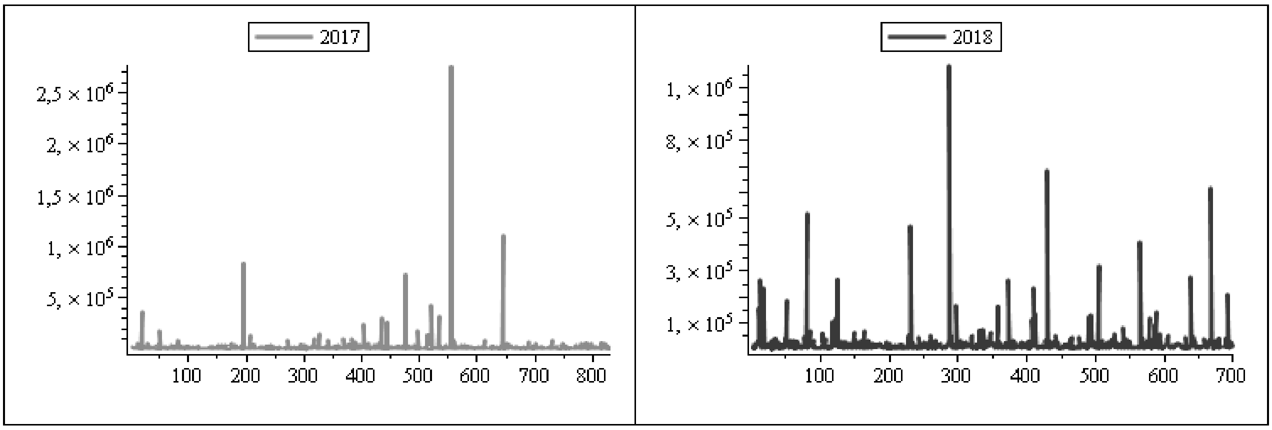

Fig. 4. **Claims Amount**

Source: Compiled by the authors.

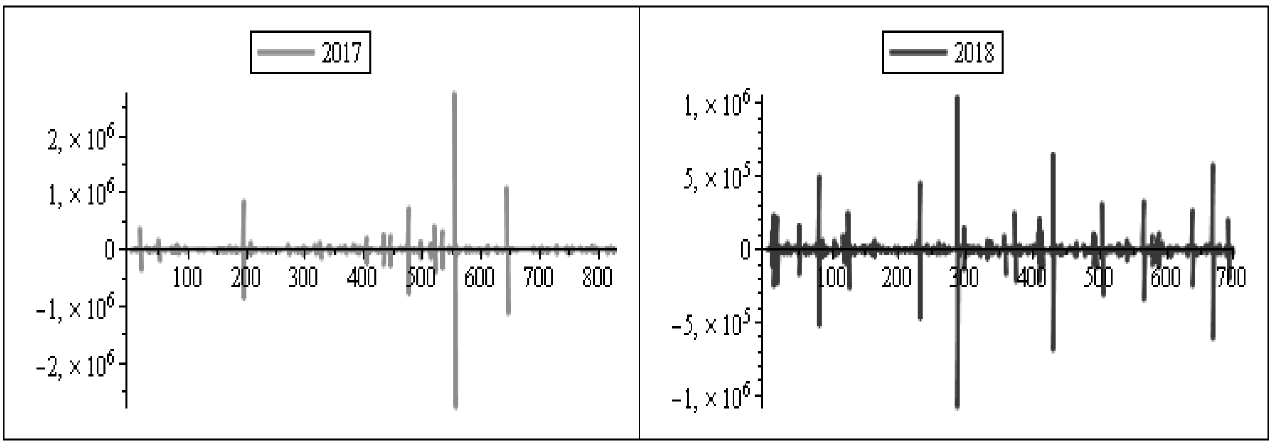

Fig. 5. **Increment of Claims Amount** $(X_{i+1}-X_i)$

Source: Compiled by the authors.

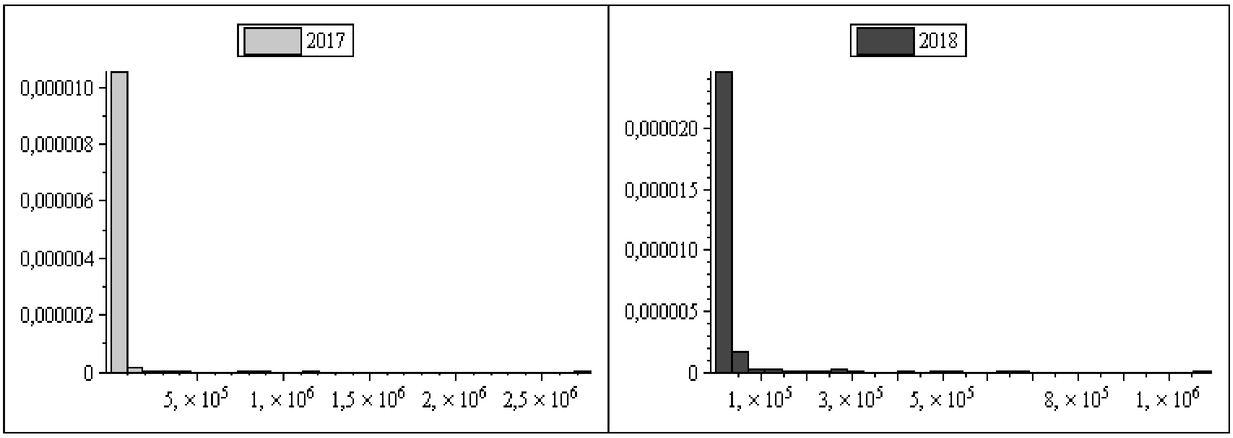

Fig. 6. **Histogram of Real Data**

Source: Compiled by the authors.

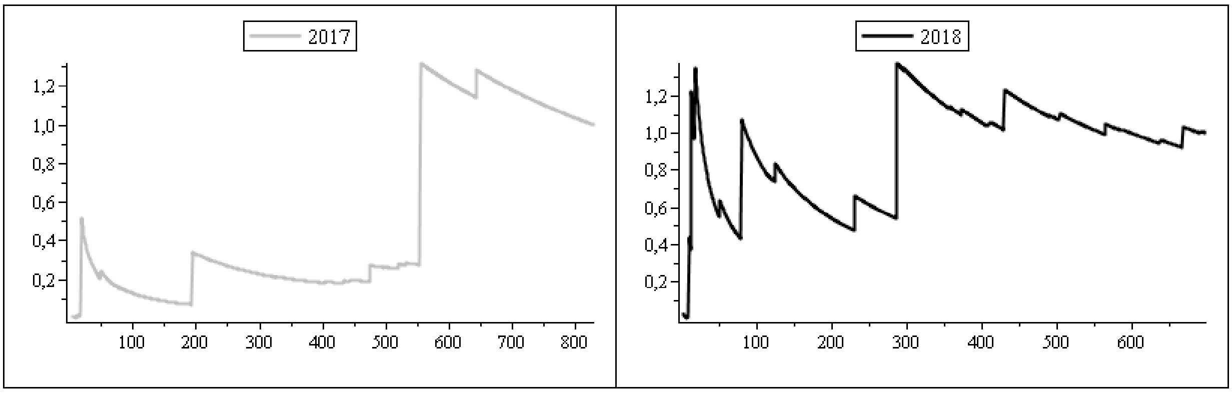

Fig. 7. **Test 1 for Real Data**

Source: Compiled by the authors.

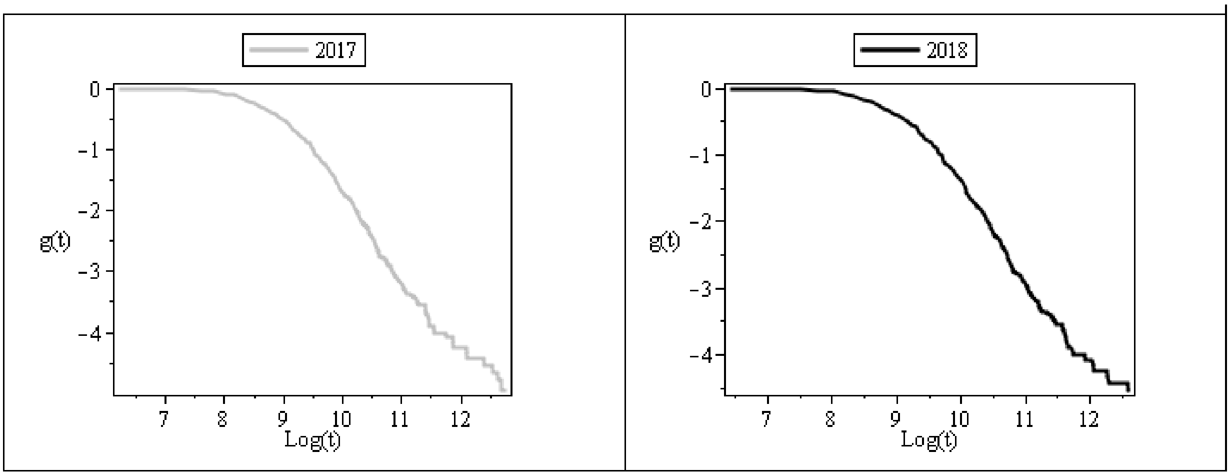

Fig. 8. **Test 2 for Real Data**

Source: Compiled by the authors.

Through *Fig. 7* and *Fig. 8*, we can see that the behavior of our data sample is closer to the stable law. The empirical variance of all our data sets (*Fig. 7*) oscillates with a high frequency, and the slope of our data sets is not finite (*Fig. 8*).

TEST AND DIAGNOSTICS

α -Stable Parameter's Estimation

There are different methods for estimation of the stable distribution parameters; we will only define the two methods used in this work.

Table 1

Parameter's Estimation of Real Data of 2018

Estimation methods	α	β	γ	μ
Koutrouvelis	1.0571	1	5.7990e+03	7.2869e+04
McCulloch	0.9563	0.9836	4.9322e+03	-6.1482e+04

Source: Compiled by the authors.

Quantile Method:

McCulloch [13] generalized the sample quantile methods for symmetric stable laws $(\beta = 0, \mu = 0)$ with $\alpha > 1$ of Fama and Roll (1971) [14] and provided consistent estimators of all four stable parameters (with the restriction $\alpha > 0.6$. He uses five sample quantiles (with q = 0.05, 0.25, 0.5, 0.75, 0.95) and matches the observed quantile spread with the exact quantile spread in stable distributions, for more details see [15] or [16].

Empirical Characteristic Function Method:

Koutrouvelis [17] presented an accurate regression-type method which starts with an initial estimate of the parameters and proceeds iteratively until some prespecified convergence criterion is satisfied. The regression method is based on empirical Characteristic Function [16].

For our estimation, we used stbl code of M. Veillette [18]. stbl is a free MATLAB library for working with alpha stable distributions. The results obtained for 2018 are summarized in the *Table 1*.

Diagnostic with Kernel Densities and Cumulative Distribution Function:

First, we use the diagnostic with Kernel densities to verify whether or not the stable fit describes the claim's amount data well. This consists in making a smoothed density plot of the real data, then comparing it to the density plots of a stable law with the parameters estimated previously. If there are clearly multiple gaps in the media, the data cannot come from a stable distribution.

In *Fig. 9* we can observe that, the claim's amount is distributed similarly to α-stable distribution and in *Fig. 10*, how give us, a comparison among the empirical cumulative distribution function CDF built

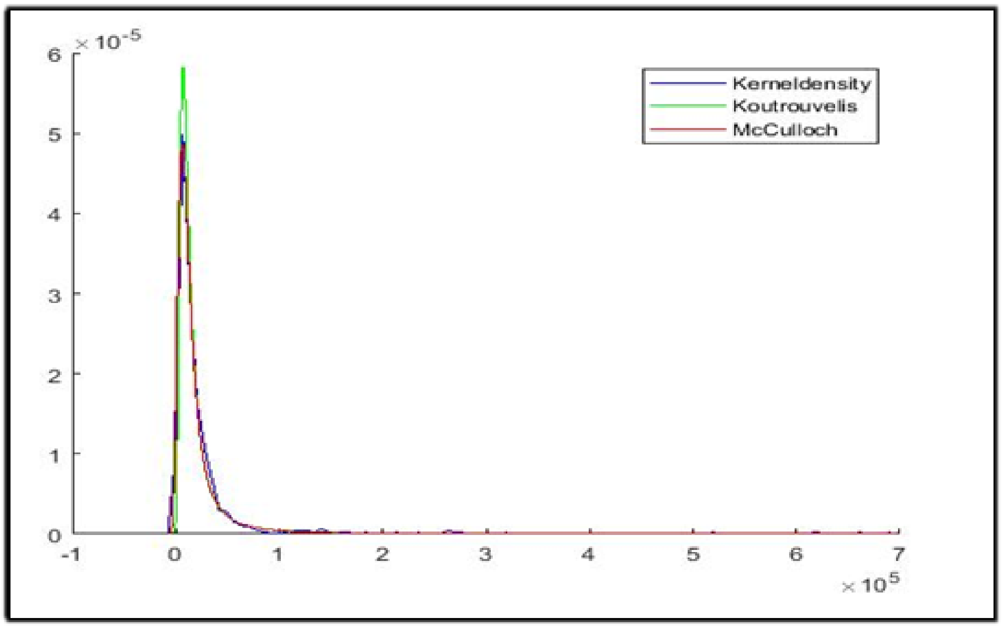

Fig. 9. **Kernel Density Estimation**

Source: Compiled by the authors.

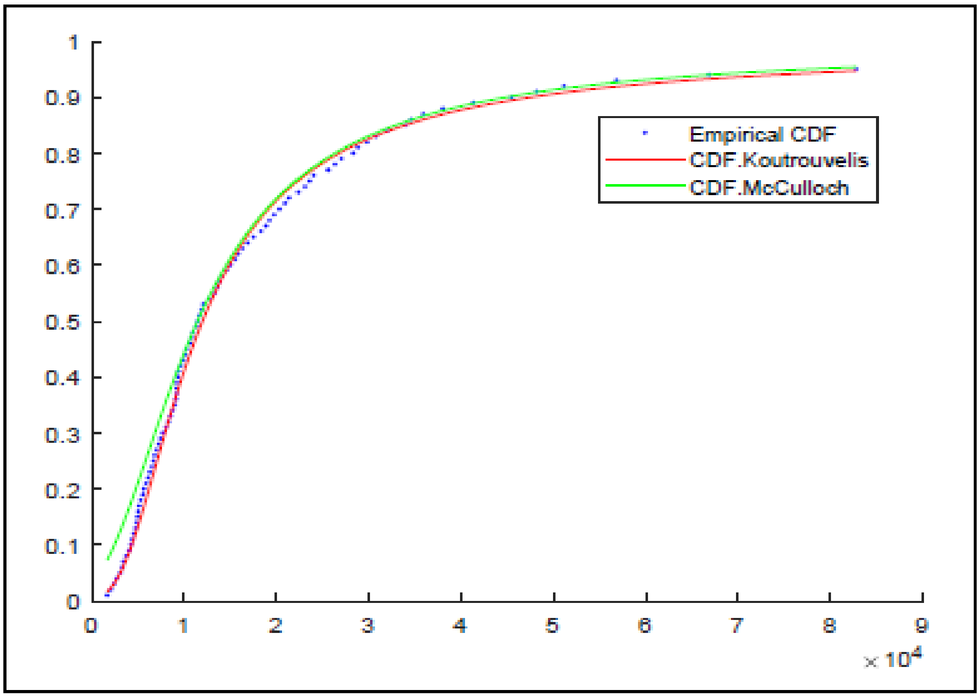

Fig. 10. **Comparison between ECDF and Stable CDF**
Source: Compiled by the authors.

from claim's amount data set and α-stable distribution CDF, we can show clearly that there are almost identical.

For confirmation of our hypothesis, we used the Kolmogorov test, how is based on the maximum distance between these curves (ECDF and CDF).

Kolmogorov Test

This test is used as a test of goodness of fit. It compares the empirical cumulative distribution function $F_n(x)$ for a variable with a CDF of specified distribution $F(x)$.

$H_0 : F_n(x) = F(x)$ Against all of the possible alternative hypotheses

$$H_1 : F_n(x) \neq F(x).$$

The null hypothesis assumes no difference between the observed and theoretical distribution and the value of test statistic 'D_K' is calculated as:

$$D_k = \text{Sup}_{x \in \text{IR}} \{ F_n(x) - F(x) \}.$$

H_0 is rejected if $D_K > D_\alpha(n)$. Where $D_\alpha(n)$ is the critical values, of the maximum absolute difference between sample $F_n(x)$ and $F(x)$.

We recall that there is no formula for the stable law distribution function, so we have estimated by using Stblcdf code of Veillette M [18].

We have $D_K < D_\alpha(n)$, so we accept H_0. Conclusion, the claim's amount can be well captured by an α-stable distribution (*Table 2*).

Table 2

The Results of Kolmogorov Test

Level of significance α Level of significance α		5%	1%
Critical values $D_\alpha(n)$		0.140	0.167
Statistic of Kolmogorov D_k	Koutrouvelis	0.031	
	Mc-Culoche	0.077	

Source: Compiled by the authors.

CONCLUSIONS

In this paper, we have reviewed the different technical diagnostics to verify and show that some data with heavy tails are well described by stable distributions, because they can model large fluctuations. We have shown through the empirical study and the diagnostics with Kernel densities that the stable distribution gives a perfect fit of the claim's amounts of car insurance; this result is very important and can help an insurer and an actuary to develop adequate strategies for risk management. In our future study, we will be interested by minimizing the ruin probability, as well as the estimation of the Lundberg coefficient for stable distributions of car claim's amount.

REFERENCES

1. Lévy P. Calcul des probabilités. Paris: Gauthier-Villars; 1925. 368 p.
2. Brookes B.C. Limit distributions for sums of independent random variables. By B.V. Gnedenko and A.N. Kolmogorov. Translated by K.L. Chung. Pp. ix, 264. $ 7.50. 1954. (Addison-Wesley, Cambridge, Mass.). *The Mathematical Gazette*. 1955;39(330):342–343. DOI: 10.2307/3608621
3. Werner T., Upper C. Time variation in the tail behaviour of bund futures returns. European Central Bank. Working Paper Series. 2002;(199). URL: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp199.pdf (accessed on 15.06.2021).
4. Bamberg G., Dorfleitner G. Fat tails and traditional capital market theory. Institut für Statistik und Mathematische Wirtschaftstheorie, Universität Augsburg. 2001;(177). URL: https://www.yumpu.com/en/document/read/36398518/fat-tails-and-traditional-capital-market-theory-gilles-daniel
5. Zolotarev V. One-dimensional stable distributions. Providence, RI: American Mathematical Society; 1986. 284 p. (Translations of Mathematical Monographs. Vol. 65). DOI: 10.1090/mmono/065
6. Nolan J.P. Univariate stable distributions: Models for heavy tailed data. Cham: Springer-Verlag; 2020. 333 p. (Springer Series in Operations Research and Financial Engineering). DOI: 10.1007/978–3–030–52915–4
7. Nolan J.P. Maximum likelihood estimation and diagnostics for stable distributions. In: Barndorff-Nielsen O.E., Resnick S.I., Mikosch T., eds. Lévy processes. Boston, MA: Birkhäuser; 2001:379–400. DOI: 10.1007/978–1–4612–0197–7_17
8. Rolski T., Schmidli H., Schmidt V., Teugels J. Stochastic processes for insurance and finance. Hoboken, NJ: John Wiley & Sons, Inc.; 1999. 680 p. (Wiley Series in Probability and Statistics).
9. d'Estampes L. Traitement statistique des processus alpha-stables: mesures de dépendance et identification des ar stables. Test séquentiels tronqués. Thèse. Toulouse: Institut national polytechnique de Toulouse; 2003. 143 p. URL: https://tel.archives-ouvertes.fr/tel-00005216 (accessed on 15.06.2021).
10. Chambers J.M., Mallows C.L., Stuck B.W. A method for simulating stable random variables. *Journal of the American Statistical Association*. 1976;71(354):340–344. DOI: 10.1080/01621459.1976.10480344
11. Janicki A., Weron A. Simulation and chaotic behavior of alpha-stable stochastic processes. New York, NY: Marcel Dekker, Inc.; 1994 376 p. (Monographs and Textbooks in Pure and Applied Mathematics. Vol. 178).
12. Cramér H. Collective risk theory: A survey of the theory from the point of view of the theory of stochastic processes. Stockholm: Nordiska bokhandeln; 1955. 92 p.
13. McCulloch J.H. Simple consistent estimators of stable distribution parameters. *Communications in Statistics — Simulation and Computation*. 1986;15(4):1109–1136. DOI: 10.1080/03610918608812563
14. Fama E.F., Roll R. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*. 1971;66(334):331–338. DOI: 10.1080/01621459.1971.10482264
15. Reiss R.-D., Thomas M. Statistical analysis of extreme values: With applications to insurance, finance, hydrology and other fields. 3rd ed. Basel: Birkhäuser; 2007. 511 p.
16. Čížek P., Härdle W.K., Weron R., eds. Statistical tools for finance and insurance. Berlin, Heidelberg: Springer-Verlag; 2011. 420 p. DOI: 10.1007/978–3–642–18062–0
17. Koutrouvelis I.A. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*. 1980;75(372):918–928. DOI: 10.1080/01621459.1980.10477573
18. Veillette M. Alpha-stable distributions for MATLAB. URL: http://math.bu.edu/people/mveillet/html/alphastablepub.html (accessed on 16.06.2021).

ABOUT THE AUTHORS / ИНФОРМАЦИЯ ОБ АВТОРАХ

Amel Laouar — PhD student in Pure and Applied Mathematics, Research Associate at Laboratory of stochastic modelization and data mining, University of science and technology Houari Boumedienne (USTHB), Algiers, Algeria
Амель Лауар — соискатель степени PhD в области теоретической и прикладной математики, научный сотрудник Лаборатории стохастического моделирования и добычи данных, Университет науки и технологии Уари Бумедьена (USTHB), Алжир, Алжир
https://orcid.org/0000-0001-9503-6230
Corresponding Author / Автор для корреспонденции:
amel.laouar@gmail.com

Kamal Boukhetala — PhD in Mathematics, Prof., Project manager and head of the actuarial risk modeling-simulation team at Laboratory of Stochastic modelization and data mining, University of science and technology Houari Boumedienne (USTHB), Algiers, Algeria
Камаль Бухетала — PhD в сфере математики, профессор, руководитель проекта и глава группы моделирования и имитации актуарных рисков в Лаборатории стохастического моделирования и добычи данных, Университет науки и технологии Уари Бумедьена (USTHB), Алжир, Алжир
https://orcid.org/0000-0003-3277-402X

Rachid Sabre — PhD in Mathematics, National Higher Institute of Agronomic, Food and Environmental Sciences, AgroSup, Dijon, France
Рашид Сабре — PhD в сфере математики, Национальный высший институт агрономических, пищевых и экологических наук, AgroSup, Дижон, Франция
https://orcid.org/0000-0002-4565-7757
rachid.sabre@agrosupdijon.fr
kboukhetala@usthb.dz

Author's declared contribution:
A. Laouar — defined the research problem, objectives of the study and developed the conceptual framework of the study, data curation, methodology, formal analysis, writing, review, and editing.
K. Boukhetala — was involved in writing (review and editing), supervision, assisted with formal analysis.
R. Sabre — assisted with formal analysis.

Заявленный вклад авторов:
А. Лауар — определил исследовательскую проблему, цели исследования и разработал концептуальную основу исследования, курировала сбор данных, методологию, формальный анализ, написание, рецензирование и редактирование.
К. Бухетала — написание (рецензирование и редактирование), проведение формального анализа.
Р. Сабре — проведение формального анализа.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.
Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

The article was submitted on 12.02.2023; revised on 10.03.2023 and accepted for publication on 28.03.2023.
The authors read and approved the final version of the manuscript.
Статья поступила в редакцию 12.02.2023; после рецензирования 10.03.2023; принята к публикации 28.03.2023.
Авторы прочитали и одобрили окончательный вариант рукописи.